Housing Price Cycle Interdependencies and Comovement:

A Markov-Switching Approach

Jeffrey P. Cohen
Professor of Finance and Kinnard Scholar in Real Estate, University of Connecticut
Research Fellow, Institute for Economic Equity, Federal Reserve Bank of St. Louis
Senior Research Fellow, Sinquefield Center, St. Louis University

Cletus C. Coughlin
Emeritus Economist, Federal Reserve Bank of St. Louis
Senior Research Fellow, Sinquefield Center, St. Louis University

Daniel Soques
Associate Professor of Economics, University of North Carolina Wilmington

June 30, 2023

Abstract: This paper uses a Markov-switching approach to examine why there is house price cycle comovement across some U.S. metropolitan areas (MSAs) but not others, and which MSAs cluster together for each of these reasons. Past studies have attributed common housing downturns in different regions as possible explanations for comovement. We explore other channels, and find some clusters based on common industry concentration (such as information technology), house price elasticity, as well as a cluster of MSAs that are desirable for retirees (in the sun belt). We find seven clusters of MSAs, where each cluster experiences idiosyncratic house price downturns, plus one distinct national house price cycle. Notably, only the housing downturn associated with the Great Recession spread across all the MSAs in our sample; all other house price downturns remained contained to a single cluster. We also identify MSA economic and geographic characteristics that correlate with housing price cluster membership, which implies comovement due to mobility of residents. In addition, while prior research has found housing and business cycles to be related closely at the national level, we find very different house price comovement and employment comovement across clusters and across MSAs.

Keywords: Housing Price Cycles, Comovement, Markov-switching, Spatial Dependence

JEL Codes: R30, C3

Andrew Spewak and Jacob Haas provided excellent research assistance. The authors benefitted from discussions with Michael Owyang and helpful referee comments. Views expressed here are the authors' alone and do not reflect the opinions of the Federal Reserve Bank of St. Louis or the Federal Reserve System.

1. Introduction

The comovement of house prices has been extensively studied in the housing literature to understand the factors behind simultaneous price movements in different groups of cities, even when cities may be geographically distant. Understanding why house price movements exhibit such comovement is crucial, as it has significant implications for policymaking and investor portfolios. Policymakers rely on a deep understanding of these movements to design effective policies that promote housing market stability, economic growth, and social welfare. Investors utilize insights into house price trends to make informed decisions about their investment portfolios, manage risk, and assess potential returns.

Previous studies have identified several mechanisms that contribute to house price comovement, including weather, climate, and available land. Some house price cycle comovement may be a result of recessions in some locations causing recessions in another region, which then may lead to lower house price growth in that other region. Regional "clusters" in some industries may contribute to regional housing market cycles.[1] Conversely, better overall health of an MSA's economy can lead to improvements in other MSAs' economic vitality, which indirectly impacts other MSAs' housing markets. Moreover, there can be other more direct mechanisms that lead to clustering across MSAs. For instance, higher house prices in some MSAs can lead residents to search for more affordable housing in nearby MSAs, leading to a direct house price growth impact in other MSAs. Alternatively, there can be cross-MSA housing market dynamics. For example, in the Northeast and in California, there is relatively high income per-capita and high house prices. Therefore, one might expect MSA average house prices in these areas to move similarly, as changes in these individuals' mobility could drive prices upward or downward simultaneously in all of these MSAs.

However, previous studies have not fully addressed the differences between national house price comovement across the entire United States compared to comovement between a subset of MSAs. Given that housing prices vary by location, we expect much heterogeneity in the links between metropolitan residential property markets and the national housing market leading to different clusters. National policies can have different effects in different MSAs and that when assessing the risks associated with a local housing market one should focus on local as opposed to only considering national dynamics. Previous studies also lack a comprehensive exploration of the mechanisms that lead to clustering across MSAs, such as residents seeking more affordable housing in nearby areas or cross-MSA housing market dynamics. Therefore, an under-explored set of questions are: what U.S. cities tend to see similar comovements in house prices? And why?

This study fills the gap in the literature and addresses the methodological shortcomings, by incorporating a novel similarity element into a multivariate Markov-switching framework that builds on existing clustering models. We consider two channels of house price cycle comovement across MSAs. The first channel we consider is common timing of house price recessions. By using Markov-switching dynamics, our study captures large movements between

[1] Although used in a different context than in our paper, the concept of regional clusters in some industries was used in Hamilton and Owyang (2012).

high house price growth phases and relatively low growth phases. Thus, the time-series clustering framework captures commonality in these regime shifts rather than short-horizon movements that are potentially noisier. Second, there may be feedback effects. In this scenario, increases in some MSAs' house price growth induces a rise in a particular "nearby" or "similar" MSA's house price growth, which in turn can cause additional house price growth in the other MSAs, etc.[2] Incorporating such feedback or multiplier effects can result in more precise estimates of the effects of variables under consideration. Our time-series clustering model that incorporates the similarity (in terms of geographic proximity) outperforms the model without one, across a number of model specifications. In a second step, we use a multinomial logistic model to investigate which characteristics of MSAs in the same cluster tend to be correlated with cluster membership.

We find seven distinct house price clusters among the top 100 U.S. MSAs. Geographic proximity is important for house price cycle comovement in some of these clusters, even after controlling for distances between MSAs. Other clusters are comprised of MSAs that are not geographically close to each other but have similar economic characteristics (such as income per-capita, the elasticity of housing supply, and house prices). In this case, at least one cluster is comprised of some MSAs that are on the opposite coasts of the U.S. This study also identifies a national housing recession that affects all clusters, while other housing downturns are specific to individual clusters. With regard to the timing of house price downturns, the Great Recession housing downturn was the only instance where there was comovement across all 100 MSAs in our sample. All other house price downturns were confined within a single cluster.

For robustness, we compare our baseline model that uses distance as the measure of MSA similarity to an alternative that gives greater weight to MSAs with similar population sizes. We find that using distance to control for similarity effects fits the data better than using population, despite both the time-series and cross-sectional variation contained in population data.

We also consider the link between house price cycles and the business cycle. To examine this issue, we apply our Markov-switching model to employment growth data in the same set of MSAs during the same time frame. We find sharp differences between the house price cycle comovement and employment comovement across cities, both in recession timing and cluster composition. We conclude that homes experience both a "volume cycle" (which previous researchers find is tightly linked with the business cycle) and a "price cycle" (which is identified in our study).[3]

Our study contributes to the existing literature by directly measuring the degree of house price comovement across groups of MSAs and providing explanations for this comovement. Our results imply that geographic proximity is important for house price cycle comovement in some clusters, even after controlling for distances between MSAs. We also highlight the differences between house price comovement and employment comovement across cities, expanding our understanding of housing and business cycles. Methodologically, our study adds to the spatial

[2] See, for example, Cohen (2010).
[3] See Leamer (2015) for discussion of the "volume cycle" and the "price cycle".

and time-series econometrics literatures by developing and estimating a generalization of other Markov-switching models.[4]

The findings of this study have important implications for various stakeholders in the housing industry, including policymakers, investors, and real estate professionals. Policymakers can use the insights gained from understanding the clustering of house price movements to design more targeted and effective policies. By recognizing the heterogeneity of housing dynamics across different MSAs, policymakers can develop region-specific strategies to address housing affordability, financial stability, and economic growth. Investors and real estate professionals can benefit from the findings by gaining a deeper understanding of the diversification potential and risk profiles of different housing markets. This knowledge can inform investment decisions, portfolio management strategies, and risk mitigation approaches in the residential real estate sector. In particular, our results may be useful to single-family housing investors in the U.S., who purchased 24 percent of all single-family homes sold in 2021, up from 15 percent per year in the prior 8 years.[5] Additionally, the methods developed in this paper will help housing researchers determine the answers to these questions at other spatial and temporal scales, both in the U.S. and internationally.

In the remainder of this paper, we first review the literature on house price diffusion in general and on Markov-switching models and their application to business cycle comovement across geographic regions. Then we present our innovation to the Hamilton and Owyang (2012) model, which allows for direct housing price growth comovement between MSAs. We next describe the data for housing price growth in the MSAs that we use in our application, and then we present results from our housing price growth Markov-switching models for the U.S., covering the period of the 1970s to 2018. The subsequent section, a crucial part of this paper, is an analysis that explores what factors determine cluster membership. We continue with a discussion of whether the house price and employment cycles are unique, and whether there is both a volume and price cycle for housing. We conclude the paper with a summary of our findings and suggestions for future research.

2. Literature Review

An important methodology that has received some (but limited) attention in the literature on house price comovement is the Markov-switching model. More generally, the Markov-switching model outlined by Hamilton (1989) is a standard framework to determine periods of expansion and recession in a time series. Hamilton and Owyang (2012) extended the Markov-switching framework to consider common recession across geographic areas in a parsimonious manner using time series clustering.

The approach of time series clustering in general was first outlined by Frühwirth-Schnatter and Kaufmann (2008). Hamilton and Owyang (2012) applied time-series clustering to state-level employment growth and found a number of sub-national business cycles underlying the national

[4] The Hamilton and Owyang (2012) Markov-switching model is a prime example of the basis of our generalization.
[5] See: https://www.pewtrusts.org/en/research-and-analysis/blogs/stateline/2022/07/22/investors-bought-a-quarter-of-homes-sold-last-year-driving-up-rents (accessed 3/4/2023).

cycle.[6] Hernández-Murillo et al. (2017) applied a similar model data on housing starts at the MSA-level. Their study found a national housing cycle that correlates with the national business cycle with deviations for three clusters of cities.

A number of recent papers use Markov-switching models for real estate applications, but not in the context of housing, include Chou and Chen (2014) and Anderson et al. (2012), who consider the relationships between Real Estate Investment Trusts (REITs) and monetary policy. Another Markov-switching study is Carstens and Freybote (2021), who analyze how the tone of REIT statements impacts the commercial real estate investing environment. Liow and Ye (2017) use switching models to study the relationships between the Global Financial Crisis and international public real estate markets. Beracha et al. (2019) focus their analysis on estimating how the commercial real estate risk premium is determined by various factors, in a Markov-switching context. Freybote and Seagraves (2018) find that investor sentiment has a positive effect on turnover in the U.S. commercial (office) real estate sector, based on estimations with Markov-switching regressions.

The literature examining the dynamics of housing prices is quite extensive. Many geographies and statistical methods are utilized in examining numerous topics. Since location characteristics are important for housing prices, the role of spatial dependencies in housing markets has been found to be of significant importance.[7] As noted by Wong, Yiu, and Chau (2012), the focus of most studies has been on correcting for bias or improving efficiency. [8] For example, Osland (2010) applied spatial econometric techniques to hedonic house price modeling in the case of privately-owned, single-family homes in Norwegian municipalities and found that such modeling added explanatory power relative to a base model. Hyun and Milcheva (2018), in the context of apartment transactions in South Korea, show that there are asymmetric spatial effects. Nearby apartment prices serve a benchmark function during a boom, but they are far less useful in capturing housing market dynamics during a bust. Finally, Clauretie and Daneshvary (2009) consider spatial aspects of foreclosures, and develop methods to control for time on the market. One housing study that relies on Markov-switching models is Nneji et al. (2013), who examine bubbles in the U.S. housing market.

We next discuss some papers dealing with house price comovement,[9] rather than house price diffusion.[10] Specifically, the analysis is on the movement of housing prices across regions contemporaneously rather on the movement of prices in one region over time in response to an

---

[6] Another related paper focused on MSAs, is by Arias et al. (2016). Based on a dynamic factor model, they highlight the heterogeneity of business cycles at the metropolitan level. In a related paper, Owyang et al. (2013) find much heterogeneity in employment cycles across 57 large U.S. cities.
[7] A voluminous literature exists with Dubin (1988) and Can (1992) are frequently cited as the first papers to apply spatial econometric techniques in the context of real estate prices. More recently, Baltagi, Fingleton, and Pirotte (2014), Baltagi and Bresson (2011), Besner (2002), and Basu and Thibodeau (1998) developed more rigorous approaches to estimate spatial econometrics in the context of hedonic housing price models.
[8] In contrast, Wong, Yiu and Chau (2012), using spatial econometric techniques, argue that the price discovery process is the economic explanation for the spatially correlated prices of Hong Kong apartments.
[9] Fischer et al. (2021) focus on comovement at a very micro level (the New York City borough of Manhattan) and find that comovement is very local over the period of 2004-2015.
[10] For a recent example of the latter, see Cohen and Zabel (2021).

initial change in price in another region. We begin by briefly discussing papers that analyze (pairwise) house-price comovement of a small number of areas via various statistical methods. Most of these papers focus on the statistical method and fit rather than the underlying economics.

Using housing price indices for four Census divisions in the Western and Midwestern United States (Pacific, Mountain, West North Central, and East North Central), Zimmer (2015) compared a Gaussian copula approach with vine copulas, a more flexible approach. He found that the latter approach produced a better data fit and much stronger correlations between housing price movements, especially during extreme price changes.

Another comovement paper is by Huang, Peng, and Yao (2019). Using housing price indices for four "Sand States" (California, Florida, Arizona, and Nevada), they review the methods used in modeling housing price comovements and then propose using a self-weighted quasi-maximum exponential likelihood estimator. They found asymmetric dependence of housing prices between certain states.

Using cointegration as well as structural estimation, Klyuev (2008) found that regional house prices across Census regions the United States became more synchronized in the early 1990s, suggesting a common national housing market expansion. His work also anticipated the major correction of housing prices as part of the Financial Crisis/Great Recession.[11]

Other research that motivates the potential reasons for comovement includes Oikarinen et al. (2018), who imply that the long-term supply elasticity of house prices varies substantially across areas. Leamer (2015) indicates that housing faces a volume but not a price cycle, because house prices are sticky downwards. In other words, homeowners will simply wait out a downturn rather than lower prices to sell. This implies that the housing supply elasticity is an important determinant of housing cycles.

Zhu et al. (2013), in their U.S. regional-level analysis, appeal to an approach by Case et al. (1993) that leverages similarity in economic variables. Zhu et al. (2013) note that "geographic closeness" (which can be proxied for by latitude and longitude of the MSA) can be important reasons for commonality in house prices. Their approach also considers similarities in house prices, employment, and income across U.S. regional housing markets as potential reasons for comovement in house prices. Kallberg et al. (2014) also mention how income (e.g., GDP) similarities can be important reasons for U.S. MSA house prices to move together. Finally, Choi and Hansz (2021) use the Wharton Land Use Regulatory Index (WRLURI) [developed by Gyourko et al. (2008)] as a control variable in determining the reasons for U.S. MSA house price comovement.

[11] Closely related to studies on comovement are studies on long-run convergence. One recent example is Barros, Gil-Alana, and Payne (2012). Using U.S. state housing price indices and overall U.S. housing prices and fractional integration and cointegration techniques, they raise doubts about long-run convergence in U.S. state housing prices and the presence of the ripple effect. On the other hand, Holmes, Otero, and Panagiotidis (2011) focus on long-run convergence across states and MSAs. Using pairwise unit root rejections, they find evidence supporting long-run convergence, with a speed of adjustment inversely related to distance.

Another paper that attempts to identify economic reasons for their statistical results, albeit not U.S. oriented, is Merikas et al. (2012).[12] In their study, Merikas et al. (2012) explore whether the comovement of housing prices across seven Eurozone countries implied convergence of their housing markets. They found that the movement of housing prices was affected by not only common fundamentals (e.g., GDP and interest rates), but also by idiosyncratic and structural factors, such as demographics, tax systems, and government interventions, which determine the duration and strength of housing cycles in these countries. In addition, they explore differences in behavior in expansions versus contractions, which is similar to our research.[13]

Turning to the analysis of (group) house-price comovement of variable clusters of areas, one finds a smaller number of papers. For example, Clark and Coggin (2009) examined the time series properties of housing prices of US census regions to assess the convergence of these prices. After reducing the number of regions to two super-regional factors, the evidence for club convergence was mixed.

Apergis and Payne (2012), using housing price indices for U.S. states and the club convergence and clustering procedures of Phillips and Sul (2007), found three convergence clubs.[14] One club consists of 29 states encompassing the BEA regions of the Mideast, New England, and Rocky Mountain plus several states from other regions. Another club consists of 19 states primarily in the Southeast and Plains regions plus states from a few other regions. The third club consists of two states in the Southeast region – Arkansas and Mississippi. The underlying factors determining these clusters, such as migration and spatial arbitrage, are not explored.

A final paper that is closer to our approach than Apergis and Payne (2012), is by Prüser and Schmidt (2021). Using a Markov-switching model and national and state-level housing prices, Prüser and Schmidt (2021) identify three house price regimes: a nationwide boom regime, a spatially limited (generally coastal) bust regime, and a nationwide bust regime. Thus, they are able to distinguish national house price cycles as well as cycles confined to a limited number of states. They focus on controlling for stochastic volatility in their econometric framework, opposed to regional spillovers in a similarity matrix. Second, they look at state-level cycles opposed to MSA-level housing cycles. Finally, Prüser and Schmidt (2021) find only one idiosyncratic regional cycle whereas others (e.g., Hernández-Murillo et al., 2017) find multiple regional cycles.

In more general contexts, similarity weights matrices can take a variety of forms, including those where each element gives equal weight to contiguous neighboring jurisdictions and zero weight to other jurisdictions. Such matrices are common in the spatial econometrics literature, as in

[12] They identified a number of other cross-country studies that have explored the impact of synchronized monetary policy, integrated financial markets, financial liberalization, and global business cycle linkages on the comovement of house prices.

[13] The role of housing in business cycles is analyzed by Álvarez et al. (2009). They found that GDP cycles among Germany, France, Italy, and Spain showed a high degree of comovement, much higher than the comovement of housing prices.

[14] See Apergis and Payne (2012) for an extensive list of references exploring convergence in regional housing markets outside the United States, frequently in the United Kingdom.

LeSage and Pace (2009). Another possibility is to allow weights to depend on the inverse distance between two jurisdictions, so that nearby MSAs (as in our case) are given greater weight than those further apart. An additional possibility, which was first proposed by Case, Rosen and Hines (1993), is to allow for "similar" jurisdictions to be given greater weight, where measures of similarity can include population size (either total population or population consisting of various minority groups), gross state product, and income, among others. While this approach allows for many alternative forms of similarity, there are potential concerns of endogeneity with some of these matrices that are not a concern with the inverse distance or contiguity approaches.

While Owyang et al. (2008) used a Markov-switching model to examine employment growth across cities, another related issue of interest has been the relationship between housing cycles and employment cycles. Past research, including Leamer (2007) and Hernández-Murillo et al. (2017), has found that business cycles (e.g., employment) and housing market cycles tend to move in tandem, particularly at the national level. Leamer (2015) similarly suggests that "Housing has a volume cycle, not a pricing cycle." However, this statement was made with respect to the housing market's link with the business cycle. Groshen and Potter (2003) and Jaimovich and Siu (2020), among others, find evidence of "jobless recoveries" where the duration of unemployment outlasts the housing market downturns. Whether there are linkages between housing and employment cycles is an important question for those who attempt to forecast housing prices, and the potential synergies in these cycles are among the issues that we explore below.

3. Approach: Clustered Housing Cycles with Comovement

Our approach has similar roots as some of the studies described above, in that we apply spatial econometric techniques to a Markov-Switching model. But we offer several additional contributions. In our case, we examine housing prices across 100 U.S. MSAs. Our focus on the importance of the stage of the business cycle in the context of spatial dependence has not been highlighted previously. An important element of our extension of Hamilton and Owyang (2012) is our incorporation of the term involving a similarity weights matrix. In addition to the similarity weighting methods, our paper includes the approach of time-series clustering. To our knowledge, similarity weighting matrices have not been incorporated in the Markov-switching literature, which makes our approach novel.

Our paper differs from Hernández-Murillo et al. (2017) in a number of ways. First, our paper focuses on house price comovement, opposed to housing starts. Second, our model captures comovement through the similarity weighting matrix, but their model imposes that there are no direct comovement across cities. Finally, our time sample begins in 1975, which allows us to capture more recessions than their sample that starts in 1989.

The methodology outlined below parallels Hamilton and Owyang (2012), with several differences. First, our dependent variable is house price growth instead of employment growth. Second, we allow for house price growth to be directly correlated across MSAs, instead of limiting ourselves to house price growth in a particular MSA to depend on other MSAs' house price growth through contemporaneous recessions. Recall equation (1) in Hamilton and Owyang (2012):

$$y_t = \mu_0 + \mu_1 \odot s_t + \varepsilon_t , \qquad (1)$$

where $y_t = (y_{t1}, \dots, y_{tN})$ is an $(N \times 1)$ vector and in our application, y_{tn} is house price growth for MSA n at time t, $s_t = (s_{t1}, \dots, s_{tN})$ is a $(N \times 1)$ and $s_{tn} = 1$ when MSA n is in recession at time t, and 0 otherwise; $\odot$ is element-by-element multiplication. μ_0 and μ_1 are the average house price growth in an expansion and recession, respectively. Also, $\varepsilon_t \sim iid(0, \Omega)$, s_t and ε_t are independent for all $t = 1, \dots, T$, and s_t follows a first-order Markov chain represented by the matrix P.[15] The variance-covariance matrix Ω is assumed to be diagonal with diagonal elements σ_n^2. This diagonality assumption ensures comovement is entirely captured by common recessions in the vector s_t and the similarity matrix discussed next.[16]

A crucial assumption is that there is no direct correlation in y_t across states (from the diagonal assumption of Ω); the only reason why y_t may be correlated across states is due to the possibility of recessions that are correlated across MSAs.[17] Now, we relax this assumption and generalize equation (1) as follows, to allow, in the context of our application, for the potential of direct house price growth correlation across MSAs:

$$y_t = \rho \boldsymbol{W} y_t + \mu_0 + \mu_1 \odot s_t + \varepsilon_t , \qquad (1')$$

where $\boldsymbol{W}$ is a symmetric, $N \times N$ similarity matrix, and ρ is a constant with $|\rho| < 1$. If we were to assume that an MSA's house price growth rises from nearby MSAs' house price growth (although in our model this relationship is not restricted to be in the positive direction), then $0 < \rho < 1$. The value of ρ in this range indicates the degree of whether or not the feedback effects rate is large (i.e., close to 1) or small (close to 0). We describe the concept of the feedback effects below.

The similarity matrix $\boldsymbol{W}$ has the (n, j) element equal to $1/d_{nj}$ if region n is a "neighbor" to region j, and 0 otherwise. Note that we can vary the definition of d_{nj}, and examine the robustness of our results to various definitions for d_{nj} . For instance, d_{nj} might be the Euclidean distance between the centroids of MSAs n and j. It could instead be the number of "neighbors" that MSA j has (where the "neighbors" could be contiguous or based on some other measure of similarity, such as the absolute value of the inverse of the difference in the population between

[15] This model assumes constant transition probabilities across time. Francis et al. (2019) allows various global shocks to influence the transition probabilities, so they are time varying. Because our study is primarily on which entities comove and not on the proximate shocks causing common downturns, we opted for the more parsimonious framework of constant transition probabilities.

[16] Francis et al. (2019) show that this assumption can be loosened by estimating the full covariance matrix using the method outlined by Carriero, Clark, and Marcellino (2019), but the full covariance model is dominated by our more parsimonious framework due to substantially fewer parameters and more intuitive cluster compositions.

[17] See Hamilton and Owyang (2012).

city n and j, or other economic or demographic variables, which we discuss in the literature review section above).

The specification in (1') implies the possibility of feedback effects, similar in spirit to Cohen (2010), as increased house price growth in some specific set of MSAs leads to increased house price growth in a neighboring MSA, which in turn impacts house price growth in the specific set of MSAs, etc. Therefore, the total effect may be somewhat larger than the effect that would be apparent without such feedback effects. It is of interest to examine how such feedback effects can impact house price growth in an MSA, compared with the situation where there is no direct interaction between the house price growth rates in different MSAs.

To demonstrate how to empirically model potential direct house price growth feedback effects, we can rewrite the above equation (1') as:

$$[\boldsymbol{I} - \rho \boldsymbol{W}]y_t = \mu_0 + \mu_1 \odot s_t + \varepsilon_t$$

$$y_t = [\mathbf{I} - \rho \boldsymbol{W}]^{-1}[\mu_0 + \mu_1 \odot s_t + \varepsilon_t]$$

$$y_t = \left[\sum_{i=0}^{\infty} \rho^{\boldsymbol{i}} \boldsymbol{W}^i\right][\mu_0 + \mu_1 \odot s_t + \varepsilon_t] \quad (1'')$$

First, we know that $\boldsymbol{W}^0 = \boldsymbol{I},$ where $\boldsymbol{I}$ is an N by N identity matrix. Note, for instance, that $\boldsymbol{W}\mu_0$ is the weighted average of the "neighboring" MSAs' average house price growth in an expansion, and $\boldsymbol{W^2}\mu_0$ is the weighted average of the second-order neighbors' average house price growth in an expansion (i.e., the weighted average of all neighbors of the neighbors' average house price growth), etc.

In addition to the feedback due to similarities across MSAs, we also incorporate time-series clustering into the model.[18] Namely, we assume there are a "small" number of clusters. For cluster 1, for example, there is a $(N \times 1)$ vector $h_1 = (h_{11}, \dots, h_{N1})'$. If MSA n is a member of cluster 1, the nth element of h_1 equals 1, and 0 otherwise. There is also an aggregate regime indicator, $z_t \in \{1, 2, \dots, K+2\}$. When $z_t = k$ for $k = 1, \dots, K$, then all MSAs that are members of cluster k are simultaneously in a house price recession while all other MSAs are in house price expansion. We call these first K regimes "idiosyncratic cluster recessions." The remaining two aggregate regimes, $K+1$ and $K+2$, are national house price recession and national house price expansion, respectively. In a national recession regime, all MSAs are in recession (i.e., h_{K+1} is a $N \times 1$ vector of ones). National expansion occurs when all MSAs are in house price expansionary phases and therefore h_{K+2} is a $N \times 1$ vector of zeros.

Now, define $\mu_n = [\mu_{n0}\ \mu_{n1}]'$ and $V(z_t, h) = [1, h_{n,zt}]'$. Then, we can rewrite (1'') as:

[18] The approach we follow below parallels Hamilton and Owyang (2012) and Frühwirth-Schnatter and Kaufmann (2008).

$$y_t = \left[\sum_{i=0}^{\infty} \rho^{\boldsymbol{i}} \boldsymbol{W}^i\right] [\mu_n' V(z_t, h) + \varepsilon_t]$$

If the values of $(h_1, \dots, h_K)$ are known, we have a standard Markov-switching model. But we need to understand the "configurations" of $(h_1, \dots, h_K)$ from the data, since the values of $(h_1, \dots, h_K)$ are not observed, but they influence the probability distribution functions of the observed y_t. Therefore, cluster membership is determined by similar movement in house price growth. Unlike Hamilton and Owyang (2012), we impose that a city can only be a member of a single cluster as in Hernández-Murillo et al. (2017), since this coincides better with the idea of economic regions.[19]

4. Estimation

We estimate the similarity-clustering model by using the Bayesian method of Gibbs sampling. The Gibbs sampler is an MCMC technique that partitions the parameters and latent variables into separate blocks so that each block can be sampled from its conditional distribution given the other blocks. It is particularly useful when sampling from the full joint posterior distribution is difficult or infeasible.

We assume values for both the data Y and the similarity matrix W are known. The parameters and latent variables are partitioned into six blocks: (i) the average regime growth parameters μ, (ii) the variance parameters σ, (iii) the coefficient on the similarity term ρ, (iv) the cluster membership indicators, H, (v) the transition matrix P, and (vi) the latent regime time-series indicator Z.

The prior distributions and Gibbs sampling steps follow closely with those of Hamilton and Owyang (2012) and Francis et al. (2019). The prior distributions are outlined in Table 1. We outline each of the steps of the Gibbs sampler in their entirety in the Appendix.

[Insert Table 1 Here]

The number of regional clusters K is a model selection issue. Hamilton and Owyang (2012) use cross-validation to compute marginal likelihoods in order to determine the optimal number of clusters. However, as mentioned by Hernández-Murillo et al. (2017), computing marginal likelihoods are prohibitively time-consuming when N is relatively large. Therefore, we use BIC as an approximation of the marginal likelihood to determine the optimal number of clusters.

To avoid label switching we make the standard assumption that the average growth rate of house prices is larger during expansion phases than during recession phases (i.e., $\mu_0 > \mu_0 + \mu_1$). This assumption is made to identify $s_t = 0$ as the "expansion" regime and $s_t = 1$ as the "recession"

[19] To maintain comparability to Hamilton and Owyang (2012), autoregressive terms of y were left out of the model. However, the model could be further generalized to allow for AR terms on the right-hand-side.

regime. Note that we do not impose a negative growth rate during recession phases, so in some cases house price recessions are characterized by relatively low but still positive average growth.

5. Data

The model requires two sets of data: (i) housing price growth (and employment growth for the subsequent analysis of employment clusters) represented in Y, and (ii) the similarity (or equivalently, weighting) matrix given by W. For housing price growth, we use the MSA-level house price index from Freddie Mac. The data are monthly and cover the time period 1975 – 2018. We seasonally-adjust each housing index using the standard X13ARIMA methodology. To smooth out monthly fluctuations, we use the quarterly average of the monthly observations.[20] The mean of the growth rates for each MSA are approximately in the range of 1% to 6% per quarter. MSAs on the higher end include Los Angeles (6.66%) and San Francisco (7.09%), with the lower end being comprised of Toledo, OH (2.7%) and Youngstown, OH (3.04%). MSAs with the highest standard deviation of house price growth include those in California, Florida, and Nevada. On the other hand, the most stable markets include those in the Midwest and South (such as Columbus, OH and Chattanooga, TN). Cohen, Coughlin and Yao (2016) provide a detailed analysis of house price trends in U.S. cities. We present conditional average growth rates and standard deviations in the next section; full sample statistics for the house price data are available from the authors upon request. The employment data are total nonfarm employment from the U.S. Bureau of Labor Statistics (BLS), and it consists of quarterly data for the 100 largest MSAs from the period 1975:1-2018:3. For MSAs in which employment data did not go back to 1975, we extrapolated the missing data by applying the appropriate state quarterly growth rate of employment from the BLS.

With MSA-level data, contiguity is not applicable because MSAs are spread out and in many cases, they do not have contiguous neighbors. For these reasons, we use the inverse of the Euclidean distance as elements of the similarity matrix, which can be expressed as below. The distances between each MSA pair are calculated using the latitude and longitude of the centroid of each MSA pair. Specifically, the (i, j) element of the similarity matrix, W, takes the form:

$$w_{ij} = \left[\frac{1}{d_{ij}}\right] / \textstyle\sum_k \left[\frac{1}{d_{ik}}\right],$$

where d_{ij} is the Euclidean distance (i.e., based on pairs of latitudes and longitudes) between any two MSAs, i and j; and w_{ii} = 0.

20

6. Results

We first address whether inclusion of the similarity matrix is necessary and if it improves model fit. Table 2 shows the BIC based on the posterior medians for the model with and without a similarity weighting matrix based on various numbers of regional clusters *K*. The model that includes the similarity weighting matrix is a large improvement over the model without a similarity weighting matrix as indicated by the smaller BIC for all possible choices for *K*. This result is perhaps unsurprising since the weighting matrix adds a large amount of information at the cost of a single parameter, ρ. The estimates for the coefficient ρ provide additional support to the necessary inclusion of the similarity weighting matrix. The median posterior value for ρ is 0.662 with a 90% highest posterior density (HPD) interval of [0.657, 0.667]. Such a high weight (relative to the theoretical maximum of 1) implies that the regional comovement captured by the similarity weighting matrix play an important role in explaining movements in MSA-level house prices.

[Insert Table 2 Here]

Table 2 also informs us of the optimal number of clusters K to use in our application. Since the model with $K = 7$ minimizes BIC, we use that specification for the remainder of the paper. Note that the optimal K is underestimated if no similarity weighting matrix is included in the model since the five-cluster model minimizes BIC across model specifications without the similarity weighting matrix. One potential explanation for this result is that geographic linkages are not controlled for directly in the model without a similarity weighting matrix. In this parsimonious setup, large geographic trends may dominate in determining the cluster relationships thus understating the "true" number of clusters.

6.1 Baseline Model Results

We find much heterogeneity across MSAs regime-specific parameter estimates. Table 3 shows the posterior median draw for each MSA's average growth rate under expansion (μ_0), average growth rate under recession ($\mu_0 + \mu_1$), and standard deviation (σ). The seven MSAs with the fastest housing-price growth rates during expansions are all located in California, including San Francisco, Los Angeles, and San Jose. The MSAs with the slowest growing house prices during expansions include Jackson, MS, Wichita, KS, and Augusta, GA.

[Insert Table 3 Here]

We do not find a strong link between average growth rates in expansion and average growth rates in recession, as the correlation between μ_0 and $\mu_0 + \mu_1$ is 0.01 across MSAs. In other words, across MSAs higher housing price growth rates in expansions provide little information about housing price growth rates in recessions.

The MSAs that have the shallowest house-price recessions (i.e., largest values for $\mu_0 + \mu_1$) include Pittsburgh, PA, Buffalo, NY, and San Jose, CA. Conversely, Lakeland, FL, Jacksonville, FL, and Chicago, IL tend to have the deepest house-price recessions (i.e., smallest values for $\mu_0 + \mu_1$).

In terms of residual volatility, the 20 most volatile MSAs mostly include those in California (San Jose, Bakersfield, Fresno, Riverside, Sacramento, Stockton, Oxnard, San Diego, Los Angeles) and Florida (Cape Coral, Palm Bay, North Port, Miami). Note that this residual volatility reflects movements in housing prices that are independent of the aggregate cycle, such as MSA-level housing booms or busts. MSAs with high residual volatility tend to also have higher mean growth rates in expansion, with a correlation of 0.58 between σ and μ_0 across MSAs. Additionally, we find a correlation of -0.22 between σ and $\mu_0 + \mu_1$, implying MSAs with relatively high idiosyncratic volatility also tend to have deeper house price recessions.

We now investigate which MSAs cluster together once we account for the contemporaneous cross-sectional similarity relationship. Figure 1 displays a choropleth map with different colors representing membership in one of the seven clusters. Each MSA is associated with the cluster for which it has a probability of membership greater than 0.5. Three of the MSAs in our sample belong to no cluster (i.e., none of their cluster membership probabilities exceeded 0.5); these include Birmingham, AL, Urban Honolulu, HI, and Jackson, MS. Cluster 1 is comprised of five MSAs in Florida as well as two in Georgia. Cluster 2 only has four MSAs, primarily located in Tennessee (Nashville, Memphis, and Chattanooga) along with Milwaukee, WI. MSAs in cluster 3 include seven from Ohio (Cincinnati, Cleveland, Columbus, Dayton, Akron, Toledo, and Youngstown) and five from North Carolina (Charlotte, Raleigh, Greensboro, Winston-Salem, and Durham). Cluster 4 contains most of the Northeast region included in our sample plus all MSAs in California (with the exception of Fresno) and the Washington D.C. area. Cluster 5 is a small cluster of six MSAs, which include two in Pennsylvania (Harrisburg and Scranton), Virginia Beach, VA, Albuquerque, NM, El Paso, TX, and Little Rock, AR. On the other hand, cluster 6 is relatively large with 24 MSAs spread across the United States, primarily in the West and West-South-Central Regions. Finally, cluster 7 is comprised of eight MSAs including five in the Midwest (Chicago, Detroit, Minneapolis, Kansas City, and Grand Rapids) and two in South Carolina (Greenville and Columbia).

[Insert Figure 1 Here]

There are two main takeaways from our model's cluster composition. First, geographic proximity still matters for the timing of housing market downturns even when distance is accounted for in the similarity weighting matrix. Therefore, regional models should include direct comovement (as we do with the similarity weighting matrix) as well as longer-term cyclical comovement (which we account for with clustered recessions). Second, other factors influence common house-price recession timing besides geographic proximity. This finding is illustrated by cluster 4, which includes two regional groups (the Northeast and California) that have short distances within groups but relatively large distances between them.

Recall that our clustered time series model provides recession timing when house prices are in (i) national expansion, (ii) national recession, or (iii) an idiosyncratic cluster recession. Figure 2 displays the posterior probability of each of these possibilities at each time period in our sample. We also indicate NBER recession dates with gray bars for comparison to the national business cycle. Table 4 displays this timing in a tabular format like how the NBER outlines historical recessions.[21] Periods of national house-price expansion are rare particularly after the recessions of the early 1980s.[22] The only period of a national recession in house prices is 2007-2011, which brackets the Great Recession. This time-period is the only one where enough MSAs were in a downturn to be deemed a national house-price recession according to our model. Besides the Great Recession, idiosyncratic cluster house-price recessions are prevalent since 1984. That is, we find that although most MSAs are in house-price expansion at any given time-period, different regions across the United States experience their own house-price downturns. Some clusters capture one-time events, such as cluster 6 picking up the downturn from 1983 until the middle of 1989 or cluster 7 picking up an idiosyncratic house price downturn in 1982.

[Insert Figure 2 Here]

[Insert Table 4 Here]

The Markov assumption about the regime variable provides some insight into how timing of the national and cluster cycles interact. Table 5 shows the posterior median for the transition probabilities for the aggregate regime variable z_t. The national expansion regime is less persistent than the other regimes with a probability of 0.33 of continuing in a national expansion at time t conditional on being in national expansion at $t-1$. The most likely transitions from a national expansion are into an idiosyncratic recession in either cluster 2 or 3. The national recession regime is relatively more persistent than the national expansion regime, and the most likely transition out of a national recession is into a localized recession for cluster 3. The clusters that are most likely to transmit a local recession to a national one are cluster 1 and cluster 7, each with a 0.17 probability.

[Insert Table 5 Here]

6.2 Comparison to Population Similarity Matrix

In our baseline specification we assume house price comovement occurs due to geographic proximity of MSAs, and therefore use inverse distance in the similarity matrix. However, the literature suggests regional comovement is captured by several alternative metrics (as in Case, Hines and Rosen, 1993), including major population centers. Thus, in this section, we consider an alternative similarity matrix using MSA-level population.

The alternative metric of population complicates our modelling relative to using distance since population varies across time. Therefore, we adjust the framework outlined in section 1 to allow for a time-varying similarity matrix, $\boldsymbol{W_t}$ at time period t with time-varying elements, $w_{t,ij}$. Population at an MSA-level is available on an annual basis, so we hold $\boldsymbol{W_t}$ constant throughout the four quarters of each respective year. Finally, we trim the housing price data in Y to end in

[21] See https://www.nber.org/cycles.html.
[22] We should note that this refers to the model's regime of "national house-price expansion" when all MSAs are jointly in an expansion state. We acknowledge the difference with the more common usage of national expansion wherein most MSAs are in expansion, but not all. This state is similar to the "idiosyncratic cluster recession" where a downturn is isolated to a specific set of MSAs.

2017 to match the availability of the population data. Each element of $\boldsymbol{W_t}$, denoted as $\boldsymbol{w_{i,j,t}}$, is given as follows:

$$\boldsymbol{w_{i,j,t}} = \frac{1}{|\boldsymbol{POP_{i,t}} - \boldsymbol{POP_{j,t}}|} \Big/ \sum_j \frac{1}{|\boldsymbol{POP_{i,t}} - \boldsymbol{POP_{j,t}}|}$$

, where $\boldsymbol{POP_{i,t}}$ is the population in MSA i in year t,

and $i = 1,2,\dots,N$, $j = 1,2,\dots,N$, and $t = 1975,1976,\dots,2017$. This implies that MSAs with similar populations as MSA i receive higher weight than MSAs with dissimilar populations, in a given year, t.

To analyze which similarity metric – distance or population - fits the data best, we focus again on BIC for each model.[23] The posterior median BIC for our baseline model using distance is 96916 whereas the BIC for the alternative model using population is 100749. The model using distance in the similarity matrix has a lower posterior median BIC, implying that geographic proximity is the better measure for capturing housing price comovement in this clustering framework. This result is noteworthy given that population varies across time, thereby providing additional time series dynamics not included in the distance model. However, the static measure of distance fits better than the time-varying measure of population. The fact that the HPD interval for the BIC of the population model, [100646, 100698], does not overlap with the HPD interval for BIC of the distance model, [96826, 97017], further increases our confidence that geographic distance is the appropriate similarity variable.

6.3 Determinants of Housing Clusters

The time series clustering framework utilized in our study grouped MSAs based on similar fluctuations (i.e., expansions and recessions) in house price indices. In this section, we investigate if the MSAs in a respective cluster have similar characteristics. In other words, we address the question: why is an MSA in the same housing cluster as some MSAs but not others? The model already controls for geographic proximity through the similarity weighting matrix, but there may be other important factors that drive MSAs to be members of a specific cluster.

We begin by defining the cluster associations $\tilde{h}_n \in \{1,2,\dots,7\}$ for each MSA, which is based on the posterior cluster membership represented in Figure 1. Let X_n be a $(Q \times 1)$ vector of MSA-level observable characteristics. Our goal is to see which of the variables in X tend to increase the probability that a general MSA would be a member of cluster k. Since $\tilde{h}_n$ is a categorical variable (with no ordering – the cluster numbers are arbitrary), we use a general multinomial logistic model that takes the following form:

$$\Pr\left(\tilde{h}_n = k | X_n\right) = \frac{\exp\left(\alpha_k + \beta_k' X_n\right)}{\sum_{j=1}^{7} \exp\left(\alpha_j + \beta_j' X_n\right)},$$

[23] Full results (e.g., parameter estimates, cluster membership, etc.) for the model with a population spatial similarity matrix are available from the authors upon request.

where $\beta_k = [\beta_{k1}, \ldots, \beta_{kQ}]$. We set the reference category to cluster 7 which implies the restrictions $\alpha_7 = 0$ and $\beta_7 = 0$.[24] In logit models, the coefficient β_{kq} represent the marginal effect of variable X_{nq} on the log odds of a MSA being in cluster k compared to being in the reference cluster (in our normalization, cluster 7).

To ease interpretation of the effect of X_{nq}, we translate the estimated coefficients into estimated marginal effects. These marginal effects are calculated based on the implied cluster probabilities of two hypothetical clusters, which we will call MSA r and MSA s. We assume these two MSA's are identical in their characteristics X_r and X_s except for one variable X_{rq} and X_{sq}. Practically, we set all characteristics besides the qth variable to their cross-section average: $X_{r,-q} = \bar{X}_{-q}$ and $X_{s,-q} = \bar{X}_{-q}$. For the qth variable, we set X_{rq} to one standard deviation above the cross-sectional average for X_{nq} (i.e., $X_{rq} = \bar{X}_{nq} + \varrho_q$, where ϱ_q is the standard deviation for X_{nq}) and conversely set X_{sq} to one standard deviation below the average ($X_{rq} = \bar{X}_{nq} - \varrho_q$). We then calculate the implied probability of membership in cluster k for each MSA given these marginal differences in one characteristic. The difference between these two implied probabilities provides us with the estimated marginal effect:

$$ME_{kq} = \Pr\left(\tilde{h}_r = k | \bar{X}_{-q}, \bar{X}_{nq} + \varrho_q\right) - \Pr\left(\tilde{h}_s = k | \bar{X}_{-q}, \bar{X}_{nq} - \varrho_q\right).$$

In simple terms, the marginal effect tells us the difference in probabilities for a MSA with a relatively high value for a characteristic compared to a MSA with a relatively low value for that same characteristic, holding other factors constant.

We consider seven MSA-level characteristics in X. These include the Wharton Residential Land Use Regulatory Index (WRLURI, from Gyourko et al. 2008), the house price elasticity (from Saiz 2010), the average log of employment between 1975- 2018, average log of per capita income between 1975-2018, average log of the house price index from 1975-2018, the latitude of the MSA centroid, and the longitude of the MSA centroid. Descriptive statistics for these variables (excluding the latitude and longitude variables) are presented in Table 6.

[Insert Table 6 Here]

Table 7 presents the estimated marginal effects of each variable on the probability of membership in each cluster. The MSA members of cluster 1 tend to be more southern (i.e., they have a low latitude). Cluster 3 MSA's tend to be more eastern (i.e., have a higher longitude, given that the longitude values are negative). Membership in cluster 4 tends to be characterized by MSAs with low house price elasticity, high income per capita, and a high house price index. MSAs in cluster 5 tend to have low employment. Cluster 6 MSAs tend to be more in the western

[24] Note that this assumption is necessary for identification and is arbitrary. Any other cluster could be the reference cluster and the results would be unchanged.

direction (i.e., have low longitude), low income per capita, and a low house price index. None of the factors describe membership in clusters 2 and 7 in a significant manner.[25]

[Insert Table 7 Here]

Several of these findings have potentially interesting underlying explanations. Anecdotally, cluster 4 contains MSAs that have a focus on information technology (Boston, New York City, San Francisco), and are centers for the arts, television, and movie production (Los Angeles and New York City). In all these cities (and predominant industries) in cluster 4, there is relatively high income per-capita and high house prices. This implies that perhaps these cities' house prices move together because their residents have similar skills and preferences for type of housing (i.e., high density, older and in some of the very largest, wealthiest cities), so downturns in per-capita income can affect all of these MSAs by a lack of desire to move to these more expensive housing markets. Several southern cities, such as Phoenix, Atlanta, Miami and other MSAs in Florida, and cities in South Carolina, are located in cluster 1, where many retirees make choices on where to live. These cluster 1 MSAs experience similar housing downturns, perhaps because decisions on when to retire can impact housing markets in all of these retirement locations at the same time. Cluster 3 consists of MSAs in Ohio, Indiana, Iowa, and the western part of North Carolina, all of which are in an area with similar longitude. House prices in cluster 3 tend to move in a similar direction since it is likely that individuals who desire to migrate out of the lower Midwest are choosing to do so at around the same time. Cluster 5 MSAs are scattered around the country; we lack a straightforward explanation for membership. Cluster 6 has MSAs, located in the west and west south-central regions, with low employment and relatively low house prices. It is likely that individuals who initially preferred this geographic area but later chose to move due to fewer employment opportunities have chosen to do so around the same time, which could increase relative housing supply around the same time in these MSAs.

6.4 Comparison to Employment-Based Clusters: Is the Housing Cycle the same as the Business Cycle?

These previous results focus on commonality in housing price movements. Previous studies suggest that movements in the housing cycle are intertwined with the economic business cycle.[26] We investigate this idea of "housing is the cycle" by comparing the clusters from our model using housing prices to the clusters implied by a similar model which uses MSA-level employment growth. Specifically, we use the log difference in employment for each MSA.

Figure 3 presents cluster membership for each MSA based on employment growth. Substantially more MSAs (10) are not members of any cluster compared to the house price results. There are some similarities between the employment clusters and the house price clusters. Employment

[25] A helpful referee suggested we try including a control for whether or not an MSA was a "coastal" city. We tried this and found the coastal variable was insignificant for all clusters except for Cluster 2. That cluster consists of several MSAs in Tennessee, which does not seem to correlate with coastal locations. Therefore, we decided not to include a control for coastal cities in the results in Table 7.
[26] See Leamer (2007, 2015) and Hernández-Murillo et al. (2017), among others.

cluster 4 includes the Northeast, most of California, and Washington D.C. as did house price cluster 4. Employment cluster 5 includes all of house price cluster 5 besides Harrisburg, PA and Scranton, PA. The MSAs of employment cluster 6 are all included in house price cluster 6. However, employment clusters 1, 2, 3, and 7 are considerably different than their house price counterparts.

[Insert Figure 3 Here]

The starkest difference between the model with house prices and the model with employment is with the timing of the national cycle. Figure 4 displays the posterior probability of each regime for the model using MSA-level employment growth that can directly be compared to Figure 2 for house prices. Firstly, the national expansion and recession regimes are much more frequent during employment cycles than during house price cycles. The national recession timing for employment growth correlates strongly with NBER recessions, which is perhaps unsurprising given the tight link between economic activity and employment. We note that the national employment regime endures well past the end of two of the most recent major recessions of the early 2000s and 2007-2009. This finding is evidence of the jobless recoveries as documented by others, as described in the literature review section above.

[Insert Figure 4 Here]

Finally, Leamer (2015) is correct in his assessment (see the literature review section of this paper) that the housing volume cycle coincides well with the business cycle. Our study suggests that in addition to a volume cycle, housing also has a price cycle that is distinct from the economic business cycle.[27]

7. Conclusion

Using data from the 100 largest MSAs across the United States during the period 1975-2018, we investigate the degree of comovement in house prices. We extend the Hamilton and Owyang (2012) model of endogenous time-series clustering with a spatial approach to account for direct common timing of housing downturns through a Markov-switching framework.

We compare endogenous clustering models that account for similarity linkages with those that did not, and we find that the former fit the data better. Thus, it is important for researchers to account for both direct similarity measures of house price movements as well as common recession timing when capturing housing cycle comovement.

We find evidence of 7 unique time series “clusters” of MSAs where house prices tend to move in tandem, with most housing price downturns appearing to be idiosyncratic to the cluster level

[27] Our finding also differs from Hernandez-Murillo et al. (2017), who conclude that the national cycle for housing starts mimics the NBER recession dates for economic activity.

rather than national.[28] Only one house price recession, the Great Recession, is widespread enough to spread to all MSAs.

Our findings contrast with the results on the aggregate cycle of housing starts from Hernández-Murillo et al. (2017), which finds a number of periods of national house price downturns that correlate strongly with the business cycle. We reinforce our result by estimating a similar model using local employment data to show aggregate house-price cycles are much more dispersed. In other words, we find differences between the house price cycle comovement and employment comovement across cities, both in recession timing and cluster composition. This might imply a need to reconsider the fourth key point made by Leamer (2015, p. 43) that "Homes experience a volume cycle, not a price cycle." We find evidence of both types of housing cycles present for U.S. MSAs.

Returning to the results of our main model, we find much heterogeneity across MSAs in terms of their average housing price growth rates under house-price expansions and under contractions. This finding is consistent with the finding of much heterogeneity of MSA business cycles by Arias et al. (2016) and employment cycles by Owyang et al. (2013). We also find that across MSAs, higher house-price growth rates in expansion provide little insight into growth rates in contractions. With respect to the composition of clusters, we find that geographic proximity matters for downturns even when distance is accounted for in the similarity matrix. However, geographic proximity is not the only factor influencing the timing of house-price recessions, as evidenced by some individual clusters encompassing disparate geographic areas. For instance, cluster 4 (consisting of New York City, Boston, Los Angeles, San Diego, San Francisco, and others on the east and west coast seaboards) are highly correlated with the house price index, and per-capita income; many of these cities are centers of high-tech and two of them (Los Angeles and New York City) are hubs for performing arts and television, which are high-paying sectors. Industry-wide downturns in these high-paying sectors are likely to hit cluster 4 cities around the same time. These common cycles across geographically dispersed cities may encourage people with similar preferences to avoid these high-priced housing markets at approximately the same time, which could be a contributing factor for these cities' house price downturns moving in tandem.

Cluster 1, which contains much of Florida, and Phoenix, AZ, is mostly in the southern parts of the U.S., and includes enticing locations for retirees who desire warmer climates. In cluster 1, these MSAs' house prices may be moving in similar directions depending on how the preferences of these types of migrants change over time. Another interesting feature of cluster 1 is that it has not had an idiosyncratic housing downturn since the late 1970's. On the other hand, parts of this cluster may have experienced housing cycles that were not due to a national house

[28] We should note a number of caveats. The first is that cluster membership is held constant throughout the entire sample. Allowing for time-varying cluster membership would capture interesting changes in house price linkages between MSAs across time. Second, our framework uses simple fixed transition probabilities. Future work could include macroeconomic (or even regional) shocks in a time-varying transition probability framework as in Francis et al. (2019) to diagnose the proximate causes of aggregate or regional downturns.

price contraction, but those MSA-level downturns were likely not strong enough to bring down the entire cluster. A cluster is only deemed "worthy" of being declared in a house price recession at any time period if a sufficient number of MSAs experience a bad downturn. This feature of the model marginalizes out relatively minor downturns and focuses on national and relatively large regional (i.e., cluster) downturns.

Information about the regime variable is useful for understanding how the timing of the national and cluster regimes interact. We find that the national expansion regime is less persistent than the other regimes, with the most likely transition into an idiosyncratic downturn in either cluster 2 or 3. The most likely transition from the national downturn regime is into a localized downturn for cluster 3. Meanwhile, clusters 1 and 7 are most likely to transmit a local downturn into a national one.

All of these findings could aid housing investors in their timing of buying and selling if they can glean a clearer understanding of which cities can be expected to experience downturns at the same time, and why. It might also be helpful for owner-occupiers who are considering migrating to a different part of the country and are hoping to understand where they can expect prices to go in those cities. Finally, housing researchers can rely on the new innovative methods developed here to address similar issues of comovement for other countries or regions of the world.

References

Álvarez, L.-J., Bulligan, G., Cabrero, A., Ferrara, L., and Stahl, H., 2009. "Housing Cycles in the Major Euro Area Countries," Banque de France Working Paper No. 269.

Anderson, R., Boney, V., Guirguis, H. 2012. The Impact of Switching Regimes and Monetary Shocks: An Empirical Analysis of REITs, Journal of Real Estate Research, 34:2, 157-182, DOI: 10.1080/10835547.2012.12091332

Arias, M. A., Gascon, C. S., and Rapach, D. E., 2016. "Metro Business Cycles," *Journal of Urban Economics*, *94*, pp. 90–108.

Apergis, N. and Payne, J. E., 2012. "Convergence in U.S. House Prices by State: Evidence from the Club Convergence and Clustering Procedure," *Letters in Spatial and Resource Sciences*, *5*(2), pp. 103–111.

Baltagi, B.H., Fingleton, B., Pirotte, A. 2014. Spatial Lag Models with Nested Random Effects: An Instrumental Variable Procedure with an Application to English House Prices. *Journal of Urban Economics*, 80, 76-86.

Baltagi, B.H., Bresson, G. 2011. Maximum Likelihood Estimation and Lagrange Multiplier Tests for Panel Seemingly Unrelated Regressions with Spatial Lag and Spatial Errors: An Application to Hedonic Housing Prices in Paris. *Journal of Urban Economics*, 69, 24-42.

Barros, C. P., Gil-Alana, L. A., and Payne, J. E., 2012. "Comovements Among U.S. State Housing Prices: Evidence from Fractional Cointegration," *Economic Modelling*, *29*(3), pp. 936–942.

Basu, S., Thibodeau, T.G. 1998. Analysis of Spatial Autocorrelation in House Prices. *Journal of Real Estate Finance and Economics*, 17(1), 61-85.

Beracha, E., Freybote, J., Lin, Z. 2019. "The Determinants of the Ex Ante Risk Premium in Commercial Real Estate", *Journal of Real Estate Research*, 41:3, 411-442, DOI: 10.22300/0896-5803.41.3.411

Besner, C. 2002. A Spatial Autoregressive Specification with a Comparable Sales Weighting Scheme. *Journal of Real Estate Research*, 24(2), 193-211.

Carriero, A., Clark, T.E., and Marcellino, M., 2019. "Large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors," *Journal of Econometrics*, *212*(1), pp.137-154.

Carstens, R., Freybote, J. 2021. "Can the Textual Tone in REIT Financial Statements Improve the Information Environment for Commercial Real Estate Investors? An Investigation", *Journal of Real Estate Research*, 43:3, 335-354, DOI: 10.1080/08965803.2021.1985923

Case, A. C., Rosen, H. S., and Hines Jr, J., 1993. "Budget Spillovers and Fiscal Policy Interdependence: Evidence from the States," *Journal of Public Economics*, *52*(3), pp. 285-307.

Choi, C. Y., & Hansz, J. A. 2021. From banking integration to housing market integration-Evidence from the comovement of US Metropolitan House Prices. *Journal of Financial Stability*, *54*, 100883.

Chou, Y., Chen, Y. 2014. “Is the Response of REIT Returns to Monetary Policy Asymmetric?” *Journal of Real Estate Research*, 36:1, 109-136, DOI: 10.1080/10835547.2014.12091381

Clauretie, T.M., Daneshvary, N. 2009. Estimating the House Foreclosure Discount Corrected for Spatial Price Interdependence and Endogeneity of Marketing Time. *Real Estate Economics*, 37(1), 43-67.

Clark, S. P. and Coggin, T. D., 2009. “Trends, Cycles and Convergence in U.S. Regional House Prices,” *Journal of Real Estate Finance and Economics*, *39*(3), pp. 264–283.

Cohen, J. P., 2010. “The Broader Effects of Transportation Infrastructure: Spatial Econometrics and Productivity Approaches,” *Transportation Research Part E: Logistics and Transportation Review*, *46*(3), pp. 317-326.

Cohen, J. P., Coughlin, C. C., and Yao, V. W., 2016. “Sales of Distressed Residential Property: What Have We Learned from Recent Research?” Federal Reserve Bank of St. Louis *Review*, *98*(3), pp. 159-188.

Cohen, J. P., and Zabel, J. 2020. Local house price diffusion. *Real Estate Economics*, *48*(3), 710-743.

Fischer, M., Füss, R. and Stehle, S., 2021. Local house price comovements. *Real Estate Economics*, *49*(S1), pp.169-198. Francis, N., Owyang, M. T. and Soques, D., 2022. “Business Cycles across Space and Time,” Journal of Money, Credit, and Banking, 54(2), pp.921-952.

Freybote, J., Seagraves, P. 2018. “The Impact of Investor Sentiment on Commercial Real Estate Market Liquidity”, *Journal of Real Estate Research*, 40:4, 597-628, DOI: 10.1080/10835547.2018.12091513

Frühwirth-Schnatter, S. and Kaufmann, S., 2008. “Model-Based Clustering of Multiple Time Series,” *Journal of Business & Economic Statistics*, *26*(1), pp. 78-89.

Groshen, E. L. and Potter, S. M., 2003. “Has Structural Change Contributed to a Jobless Recovery?” Federal Reserve Bank of New York *Current Issues in Economics and Finance*, *9*(8), pp. 1-7.

Gyourko, J., Saiz, A., and Summers, A., (2008). “A New Measure of the Local Regulatory Environment for Housing Markets: The Wharton Residential Land Use Regulatory Index,” *Urban Studies*, 45(3), pp. 693–729.

Hamilton, J. D., 1989. “A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle,” *Econometrica*, *57*(2), pp. 357-384.

Hamilton, J. D. and Owyang, M. T., 2012. “The Propagation of Regional Recessions,” *Review of Economics and Statistics*, *94*(4), pp. 935-947.

Hernández-Murillo, R., Owyang, M. T. and Rubio, M., 2017. "Clustered Housing Cycles," *Regional Science and Urban Economics*, *66*, pp. 185-197.

Holmes, M. J., Otero, J., and Panagiotidis, T., 2011. "Investigating Regional House Price Convergence in the United States: Evidence from a Pair-wise Approach," *Economic Modelling*, *28*(6), pp. 2369–2376.

Huang, H., Peng, L., and Yao, V. W., 2019. "Comovements and Asymmetric Tail Dependence in State Housing Prices in the USA: A Nonparametric Approach," *Journal of Applied Econometrics*, *34*(5), 843–849.

Hyun, D., Milcheva, S. 2018. Spatial dependence in apartment transaction prices during boom and bust. *Regional Science and Urban Economics*, 68, 36-45.

Jaimovich, N. and Siu, H. E., 2020. "Job Polarization and Jobless Recoveries," *Review of Economics and Statistics*, *102*(1), pp. 129-147.

Kallberg, J. G., Liu, C. H., & Pasquariello, P. 2014. On the price comovement of US residential real estate markets. *Real Estate Economics*, *42*(1), 71-108.

Klyuev, V., 2008. "What Goes Up Must Come Down? House Price Dynamics in the United States," IMF Working Paper No. 08/187.

Leamer, E. E., 2007. "Housing Is the Business Cycle," Proceedings - Economic Policy Symposium - Jackson Hole, Federal Reserve Bank of Kansas City, pp. 149-233.

Leamer, E. E., 2015. "Housing Really Is the Business Cycle: What Survives the Lessons of 2008–09?" *Journal of Money, Credit and Banking*, 47(S1), pp. 43-50.

LeSage, J., and Pace, R. K., 2009. *Introduction to Spatial Econometrics*. Chapman and Hall/CRC.

Liow, K.H., Ye, Q. 2017. "Switching Regime Beta Analysis of Global Financial Crisis: Evidence from International Public Real Estate Markets", *Journal of Real Estate Research*, 39:1, 127-164, DOI: 10.1080/10835547.2017.12091462

Merikas, A., Merika, A., Laopodis, N., and Triantafyllou, A., 2012. "House Price Comovements in the Eurozone Economies," *European Research Studies*, 15(1), pp. 71–98.

Ogonna Nneji, Chris Brooks & Charles Ward. 2013. "Intrinsic and Rational Speculative Bubbles in the U.S. Housing Market: 1960-2011", *Journal of Real Estate Research*, 35:2, 121-152, DOI: 10.1080/10835547.2013.12091360

Oikarinen, E., Bourassa, S.C., Hoesli, M., Engblom, J. 2018. U.S. Metropolitan House Price Dynamics. *Journal of Urban Economics*, 105, 54-69.

Osland, L. 2010. An Application of Spatial Econometrics in Relation to Hedonic House Price Modeling. *Journal of Real Estate Research*, 32(3), 289-320.

Owyang, M. T., Piger, J.,and Wall, H. J., 2013. "Discordant City Employment Cycles," *Regional Science and Urban Economics*, 43(2), pp.367-384.

Owyang, M. T., Piger, J., Wall, H. J., and Wheeler, C. H., 2008. “The Economic Performance of Cities: A Markov-switching Approach,” *Journal of Urban Economics*, 64(3), pp. 538-550.

Owyang, M. T., Rapach, D. E., and Wall, H. J., 2009. “States and the Business Cycle,” *Journal of Urban Economics*, 65(2), pp. 181-194.

Phillips, P. C. B. and Sul, D., 2007. “Transition Modeling and Econometric Convergence Tests,” *Econometrica*, *75*(6), pp. 1771–1855.

Prüser, J. and Schmidt, T., 2021. “Regional Composition of National House Price Cycles in the US,” forthcoming in *Regional Science and Urban Economics*.

Saiz, A., 2010. “The Geographic Determinants of Housing Supply,” *The Quarterly Journal of Economics*, *125*(3), pp. 1253-1296.

Wong, S.K., Yiu, C.Y., Chau, K.W. 2013. Trading Volume-Induced Spatial Autocorrelation in Real Estate Prices. *Journal of Real Estate Finance and Economics*, 46, 596-608.

Zhu, B., Füss, R., & Rottke, N. B. 2013. Spatial linkages in returns and volatilities among US regional housing markets. *Real Estate Economics*, *41*(1), 29-64.

Zimmer, D. M., 2015. “Analyzing Comovements in Housing Prices using Vine Copulas,” *Economic Inquiry*, *53*(2), pp. 1156–1169.

Table 1: Prior Distributions

Parameter	*Prior Distribution*	*Hyperparameters*
μ_n	$N(m_{n0}, \sigma_n^2 M_{n0})$	$m_{n0} = [1, -2]', M_{n0} = I_2 \;\; \forall n$
σ_n^{-2}	$\Gamma\left(\frac{\upsilon_0}{2}, \frac{\tau_0}{2}\right)$	$\upsilon_0 = 0, \tau_0 = 0$
ρ	$N(r_0, R_0)$	$r_0 = 0, R_0 = 1$
P_i	$D(p_{1i}, \dots, p_{K+2i})$	$p_{ji} = 1 \; \forall j$

This table shows the prior distributions for the baseline model. μ_n is the vector of average house price growth rate parameters for MSA n, σ_n is the standard deviation of the shock to MSA n's house price growth, ρ is the weighting on the similarity matrix, and P is the aggregate regime transition matrix.

Table 2: BIC Model Comparison

K	*No Similarity Weighting*	*Similarity Weighting*
2	107884.37	98819.81
3	107247.93	98573.98
4	107101.29	98479.84
5	**107021.41**	98433.06
6	107084.61	98413.17
7	107125.83	**98370.62**
8	107164.86	98463.18
9	109708.73	98703.82
10	108075.02	99217.13

This table shows the model comparison using Bayesian Information Criterion for various numbers of total clusters K. The first column shows the model fit across different K for the model with no similarity matrix which is identical to the framework of Hamilton and Owyang (2012). The second column shows the model fit across different K for our model that includes a similarity matrix.

Table 3: Regime-Specific Growth Rates and Variance Parameters: This table shows the posterior mean of the MSA-specific model parameters. μ_0 and $\mu_0 + \mu_1$ are the average house price growth parameters for MSA n in expansion and recession, respectively. σ_n is the standard deviation of the shock to MSA n's house price growth.

Pop. Rank	Name	Abbr.	μ_0	$\mu_0 + \mu_1$	σ
1	New York-Newark-Jersey City, NY-NJ-PA	NYT	4.14	-0.58	4.72
2	Los Angeles-Long Beach-Anaheim, CA	LNA	5.38	-1.13	5.53
3	Chicago-Naperville-Elgin, IL-IN-WI	CHI	2.85	-3.34	3.53
4	Dallas-Fort Worth-Arlington, TX	DFW	2.70	0.12	3.59
5	Houston-The Woodlands-Sugar Land, TX	HTN	2.95	-0.88	4.01
6	Washington-Arlington-Alexandria, DC-VA-MD-WV	WSH	4.34	-0.83	4.39
7	Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	PCW	3.46	-0.69	3.22
8	Miami-Fort Lauderdale-West Palm Beach, FL	MIM	3.25	-1.25	5.87
9	Atlanta-Sandy Springs-Roswell, GA	ATL	2.72	-2.97	3.62
10	Boston-Cambridge-Newton, MA-NH	BOS	4.42	0.13	5.09
11	San Francisco-Oakland-Hayward, CA	SFC	5.56	-0.42	5.40
12	Phoenix-Mesa-Scottsdale, AZ	PHX	2.90	-1.26	6.88
13	Riverside-San Bernardino-Ontario, CA	RSB	4.69	-1.32	6.25
14	Detroit-Warren-Dearborn, MI	DWL	2.47	-2.60	5.83
15	Seattle-Tacoma-Bellevue, WA	STW	4.21	-0.69	5.75
16	Minneapolis-St. Paul-Bloomington, MN-WI	MSP	2.89	-1.44	3.92
17	San Diego-Carlsbad, CA	SDI	4.65	-0.62	5.76
18	Tampa-St. Petersburg-Clearwater, FL	TMA	3.19	-1.95	5.15
19	St. Louis, MO-IL	STL	2.14	-1.16	2.89
20	Denver-Aurora-Lakewood, CO	DNV	4.26	-0.12	3.99
21	Baltimore-Columbia-Towson, MD	BTM	3.59	-0.67	3.86
22	Charlotte-Concord-Gastonia, NC-SC	CGR	3.16	-0.29	2.56
23	Orlando-Kissimmee-Sanford, FL	ORL	2.98	-3.18	4.92
24	Portland-Vancouver-Hillsboro, OR-WA	POR	4.11	-0.68	5.04
25	San Antonio-New Braunfels, TX	SAT	2.45	-0.38	3.47
26	Pittsburgh, PA	PIT	2.21	0.96	2.79
27	Sacramento-Roseville-Arden-Arcade, CA	SYO	4.28	-1.34	6.13
28	Cincinnati, OH-KY-IN	CTI	2.38	-0.46	2.15
29	Las Vegas-Henderson-Paradise, NV	LSV	3.17	-1.56	7.11
30	Kansas City, MO-KS	KNC	2.10	-1.08	2.51
31	Cleveland-Elyria, OH	CVL	2.30	-1.48	2.90
32	Columbus, OH	COL	2.63	0.40	2.14
33	Austin-Round Rock, TX	AUS	3.61	0.69	5.12
34	Indianapolis-Carmel-Anderson, IN	IND	2.27	-0.24	2.69
35	San Jose-Sunnyvale-Santa Clara, CA	SSC	5.05	0.91	7.21
36	Nashville-Davidson-Murfreesboro-Franklin, TN	NVL	2.79	-0.09	2.96
37	Virginia Beach-Norfolk-Newport News, VA-NC MS	NFK	2.85	-0.57	4.16
38	Providence-Warwick, RI-MA	PRI	4.08	-1.36	4.86
39	Milwaukee-Waukesha-West Allis, WI	MWK	2.52	-0.95	3.69

Table 3: Regime-Specific Growth Rates and Variance Parameters: This table shows the posterior mean of the MSA-specific model parameters. μ_0 and $\mu_0 + \mu_1$are the average house price growth parameters for MSA n in expansion and recession, respectively. σ_n is the standard deviation of the shock to MSA n's house price growth.

Pop. Rank	Name	Abbr.	μ_0	$\mu_0 + \mu_1$	σ
40	Jacksonville, FL	JAX	2.94	-3.37	4.04
41	Oklahoma City, OK	OKC	2.44	-1.67	4.05
42	Memphis, TN-MS-AR	MPH	1.63	-1.53	3.10
43	Louisville/Jefferson County, KY-IN	LOI	2.49	0.34	2.09
44	Raleigh, NC	RCY	3.13	-0.01	2.93
45	Richmond, VA	RCP	2.48	-1.31	2.74
46	New Orleans-Metairie, LA	NOR	2.74	-1.42	3.72
47	Hartford-West Hartford-East Hartford, CT	HTF	2.65	-1.47	4.60
48	Salt Lake City, UT	SLC	3.77	-0.92	4.24
49	Birmingham-Hoover, AL	BIR	1.85	-1.00	2.58
50	Buffalo-Cheektowaga-Niagara Falls, NY	BUF	2.17	0.93	3.39
51	Rochester, NY	ROH	1.84	0.19	2.66
52	Grand Rapids-Wyoming, MI	GRR	2.15	-1.88	4.30
53	Tucson, AZ	TUC	2.23	-3.28	5.17
54	Urban Honolulu, HI	URH	2.88	0.78	9.92
55	Tulsa, OK	TUL	2.16	-1.31	3.63
56	Fresno, CA	FRE	3.01	-1.52	6.26
57	Bridgeport-Stamford-Norwalk, CT	BRG	3.05	-0.93	5.39
58	Worcester, MA-CT	WST	4.14	-1.17	4.44
59	Omaha-Council Bluffs, NE-IA	OMA	2.17	0.32	2.80
60	Albuquerque, NM	ABQ	2.67	-0.81	3.76
61	Albany-Schenectady-Troy, NY	ALB	2.30	-0.32	4.55
62	Bakersfield, CA	BAK	2.97	-1.28	6.30
63	Greenville-Anderson-Mauldin, SC	GNV	2.16	-0.40	2.93
64	New Haven-Milford, CT	NHM	2.69	-1.67	4.92
65	Knoxville, TN	KNX	2.20	0.24	2.73
66	Oxnard-Thousand Oaks-Ventura, CA	VEN	5.04	-1.05	5.88
67	McAllen-Edinburg-Mission, TX	MCL	1.55	-1.54	3.15
68	El Paso, TX	ELP	1.68	-0.44	3.66
69	Allentown-Bethlehem-Easton, PA-NJ	ALL	2.85	-1.30	3.84
70	Baton Rouge, LA	BTR	2.81	-1.33	4.03
71	Columbia, SC	CBA	1.55	-1.21	2.65
72	Dayton, OH	DYT	2.05	-1.13	2.77
73	North Port-Sarasota-Bradenton, FL	SAR	3.53	-0.86	6.08
74	Greensboro-High Point, NC	GNS	2.09	-0.68	2.16
75	Charleston-North Charleston, SC	CRL	3.89	-0.71	4.08
76	Little Rock-North Little Rock-Conway, AR	LRS	1.69	0.13	3.37
77	Stockton-Lodi, CA	STO	4.49	-1.59	5.99
78	Akron, OH	AKR	2.16	-1.00	2.42

Table 3: Regime-Specific Growth Rates and Variance Parameters: This table shows the posterior mean of the MSA-specific model parameters. μ_0 and $\mu_0 + \mu_1$are the average house price growth parameters for MSA n in expansion and recession, respectively. σ_n is the standard deviation of the shock to MSA n's house price growth.

Pop. Rank	Name	Abbr.	μ_0	$\mu_0 + \mu_1$	σ
79	Cape Coral-Fort Myers, FL	FTM	2.94	-1.14	6.80
80	Colorado Springs, CO	CLR	2.84	-0.50	3.64
81	Boise City, ID	BOI	3.04	-1.88	5.94
82	Syracuse, NY	SYR	1.98	-0.41	3.13
83	Winston-Salem, NC	WSA	1.86	-0.56	2.11
84	Lakeland-Winter Haven, FL	LWH	2.21	-4.07	5.04
85	Wichita, KS	WIC	1.37	-0.02	3.25
86	Ogden-Clearfield, UT	OCR	2.96	-0.75	3.97
87	Madison, WI	MDS	3.21	0.10	3.59
88	Springfield, MA	SPD	3.78	-0.91	3.78
89	Des Moines-West Des Moines, IA	DEM	2.27	-0.08	3.09
90	Deltona-Daytona Beach-Ormond Beach, FL	DDO	3.36	-1.66	5.34
91	Toledo, OH	TOL	1.46	-1.24	2.73
92	Augusta-Richmond County, GA-SC	AUG	1.44	-1.40	2.69
93	Provo-Orem, UT	PRV	3.24	-0.94	4.54
94	Jackson, MS	JAS	1.05	-0.52	4.05
95	Palm Bay-Melbourne-Titusville, FL	MEL	3.34	-1.63	6.49
96	Harrisburg-Carlisle, PA	HAR	2.19	0.06	2.15
97	Scranton-Wilkes-Barre-Hazleton, PA	SWB	2.30	-0.27	3.35
98	Durham-Chapel Hill, NC	RAD	3.02	0.69	2.45
99	Youngstown-Warren-Boardman, OH-PA	YNG	1.77	-0.83	2.63
100	Chattanooga, TN-GA	CHT	1.89	0.29	2.12

Table 4: Timing of Housing Cycle Phases

Start of New Phase	Type of Phase	Duration (in Quarters)
4/1/1975	National Expansion	1
7/1/1975	Cluster 1 Recession	1
1/1/1976	National Expansion	1
4/1/1976	Cluster 2 Recession	2
10/1/1976	National Expansion	1
1/1/1977	Cluster 1 Recession	1
4/1/1977	National Expansion	7
1/1/1979	Cluster 2 Recession	1
4/1/1979	National Expansion	3
1/1/1980	Cluster 3 Recession	2
7/1/1980	National Expansion	2
1/1/1981	Cluster 3 Recession	4
1/1/1982	National Expansion	1
4/1/1982	Cluster 7 Recession	3
1/1/1983	National Expansion	1
4/1/1983	Cluster 6 Recession	25
7/1/1989	National Expansion	2
1/1/1990	Cluster 4 Recession	31
10/1/1997	National Expansion	1
1/1/1998	Cluster 5 Recession	9
4/1/2000	National Expansion	1
7/1/2000	Cluster 3 Recession	6
1/1/2002	National Expansion	1
4/1/2002	Cluster 3 Recession	15
1/1/2006	National Expansion	1
4/1/2006	Cluster 4 Recession	4
4/1/2007	National Recession	20
4/1/2012	Cluster 3 Recession	2
10/1/2012	Cluster 5 Recession	24

This table shows the estimated aggregate regime timing for our baseline model of housing cycles. The first column provides the first quarter of the new phase, the second column shows the type of phase that is initialized, and the last column shows its duration.

Table 5: Transition Probabilities

		Previous Regime								
		Nat'l Exp.	Nat'l Rec.	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7
Current Regime	Nat'l Exp.	0.33	0.03	0.48	0.49	0.18	0.05	0.06	0.07	0.34
	Nat'l Rec.	0.03	0.69	0.17	0.15	0.03	0.05	0.03	0.04	0.17
	Cluster 1	0.09	0.03	0.35	-	-	-	-	-	-
	Cluster 2	0.10	0.03	-	0.36	-	-	-	-	-
	Cluster 3	0.15	0.06	-	-	0.78	-	-	-	-
	Cluster 4	0.09	0.03	-	-	-	0.89	-	-	-
	Cluster 5	0.09	0.04	-	-	-	-	0.91	-	-
	Cluster 6	0.06	0.04	-	-	-	-	-	0.89	-
	Cluster 7	0.06	0.03	-	-	-	-	-	-	0.49

This table shows the estimated transition matrix for the aggregate regime variable z_t. The columns show the regime transition from while the rows show which regime is being transitioned to. Note that '-' indicates that we restrict transitions between idiosyncratic cluster recessions, as in Hamilton and Owyang (2012).

Table 6: Descriptive Statistics for Cluster Determinants

	Mean	Median	Std. Dev.	Min	Max
WRLURI	0.083	0.035	0.677	-1.239	1.892
House price elasticity	1.859	1.638	0.92	0.627	5.453
Employment	749.396	387.8	1100.178	94.833	8159.833
Income Per Capita	19587.518	19333.363	3214.785	9518.871	33701.206
House Price Index	71.7	70.063	12.25	43.153	100.153

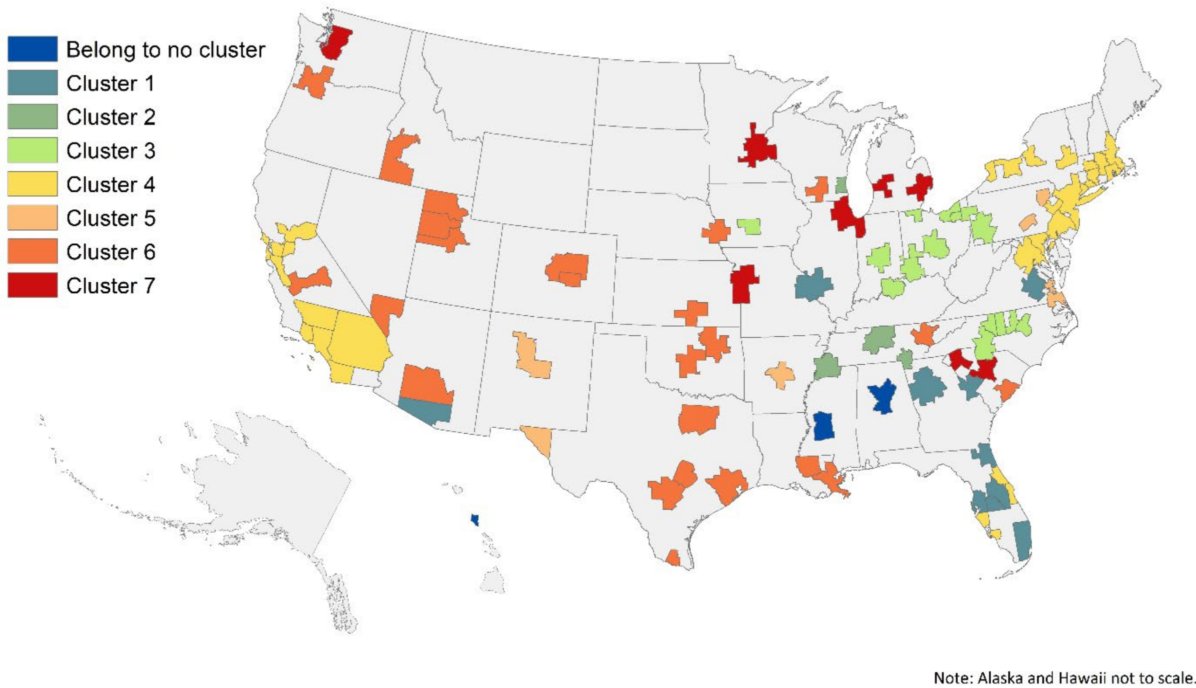
This table shows the descriptive statistics for each covariate used in the multinomial logistic regression models. The Wharton Land Use Regulatory Index (WRLURI) and the measure of house price elasticity come from Saiz (2010). Employment and income per capita are logged and averaged over their values between 1975 – 2018.

Table 7: Marginal Effects of Cluster Determinants

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7
WRLURI	0.202	0.146	-0.01	0.157	0.08	-0.205	-0.37
House price elasticity	0.202	0.014	0.048	**-0.544*** **	-0.008	0.152	0.137
Latitude	**-0.508** **	0.05	0.034	0.028	0.159	-0.178	0.415
Longitude	0.067	0.03	**0.714*** **	-0.019	-0.036	**-0.662*** **	-0.096
Employment	0.328	-0.007	-0.007	0.048	**-0.897*** **	0.207	0.327
Income Per Capita	-0.239	-0.055	0.03	**0.492* **	0.19	**-0.345* **	-0.072
House Price Index	-0.084	-0.152	-0.025	**0.823*** **	0.08	**-0.425*** **	-0.216

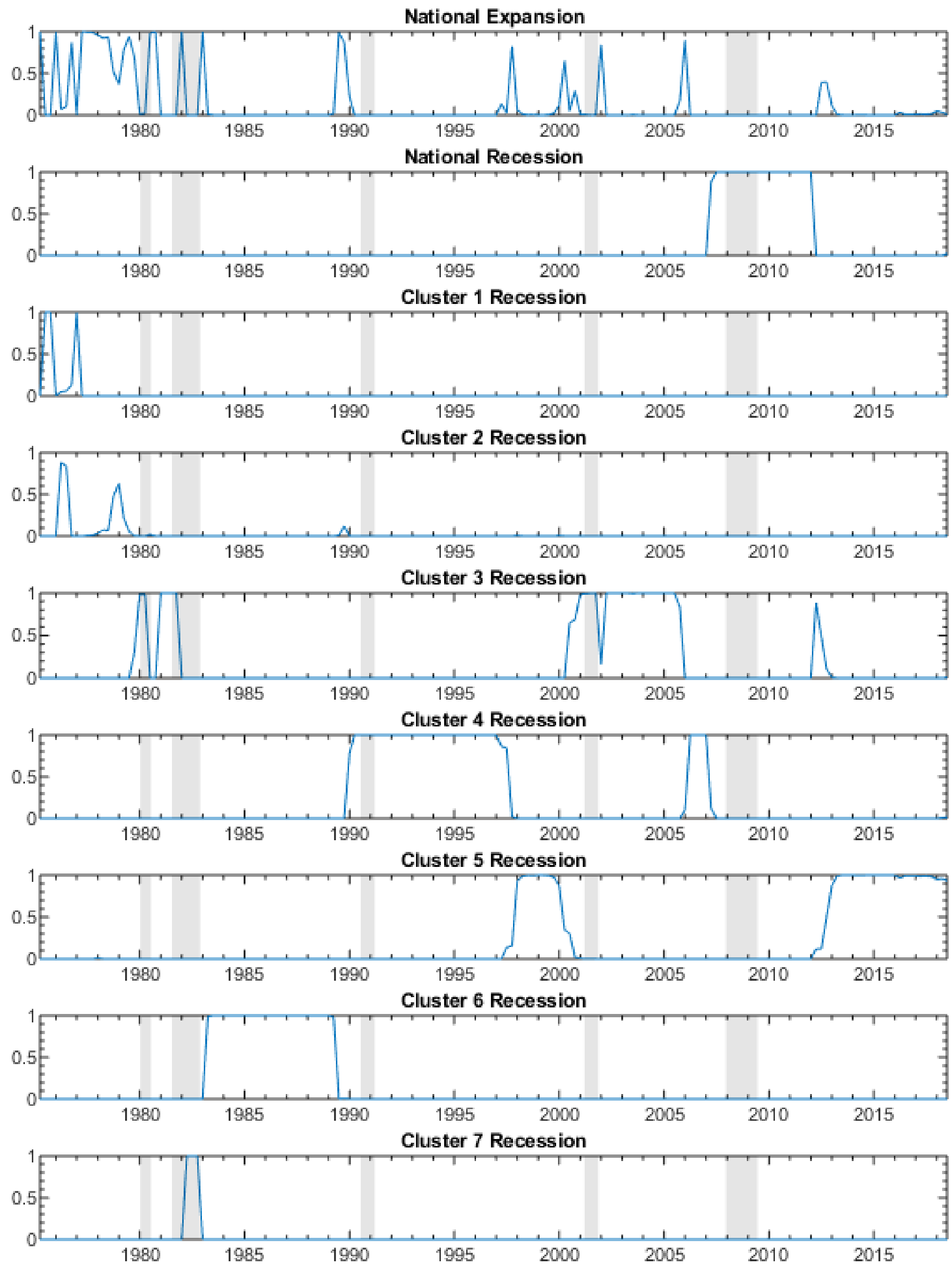
This table shows the estimated marginal effects of each covariate from a multinomial logistic regression model. The dependent variable is the cluster membership indicators obtained from our time-series clustering model that includes a similarity matrix. The Wharton Land Use Regulatory Index (WRLURI) and the measure of house price elasticity come from Saiz (2010). Latitude and longitude are based on the respective MSA centroid. Employment and income per capita are logged and averaged over their values between 1975 – 2018. *** $p < 0.01$; ** $p < 0.05$, * $p < 0.10$

Figure 1: Cluster Membership Based on House Prices



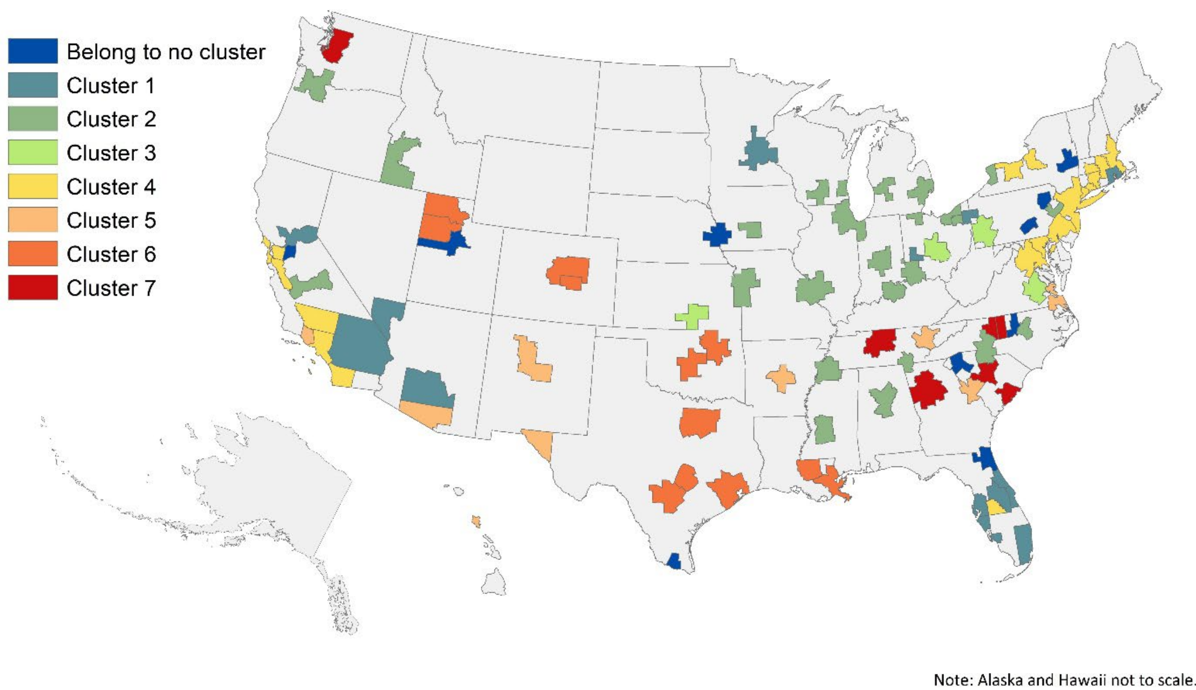
This figure shows the posterior probability of cluster membership for each MSA in our sample using our baseline time-series clustering model with a similarity matrix. In this framework, MSAs cluster together based on similarities in house price fluctuations. Gray areas are not included in our sample.

Figure 2: Timing of House Price Downturns



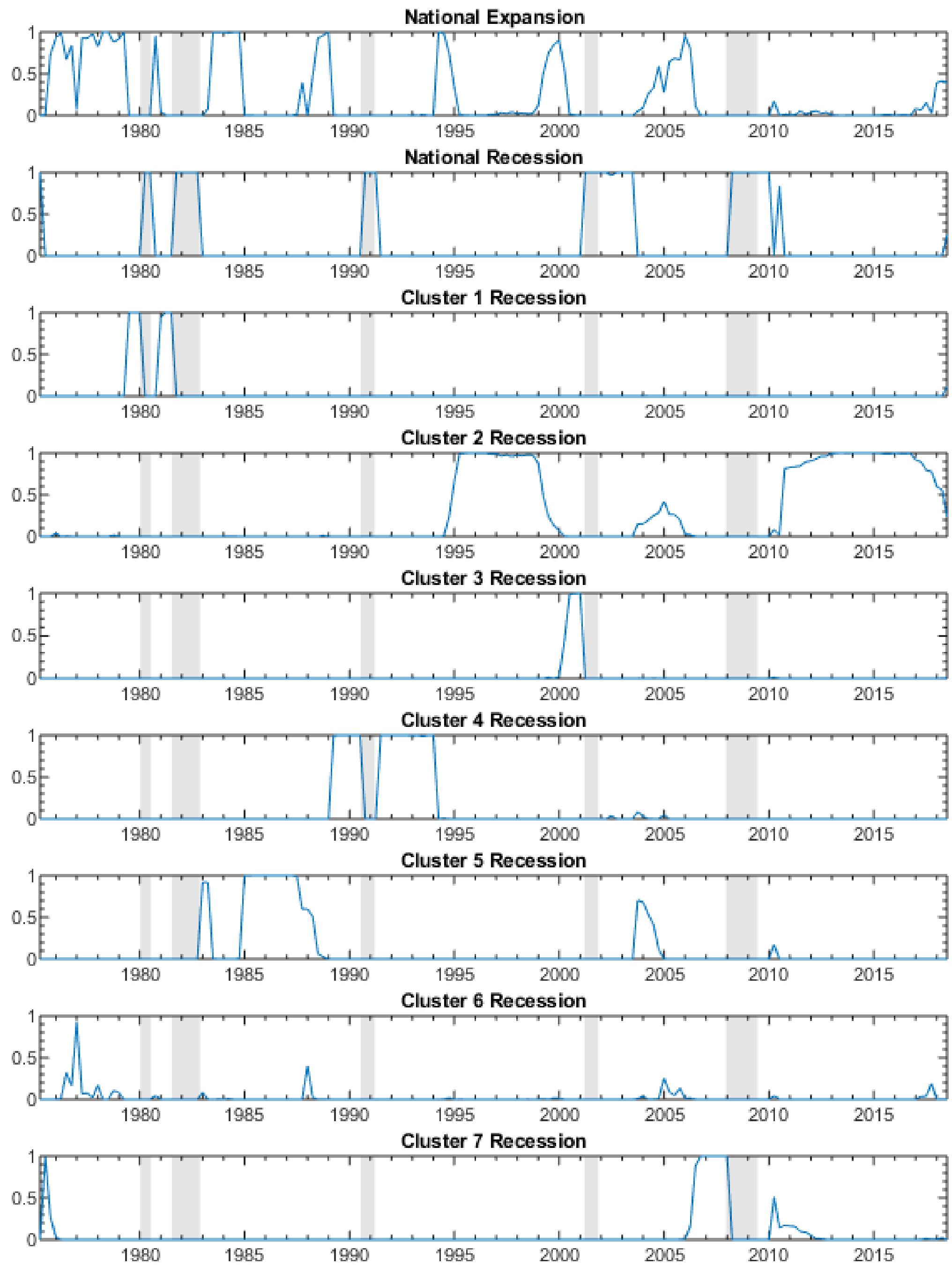
This figure shows the posterior probability of regime timing for the aggregate variable z_t using our baseline time-series clustering model with a similarity matrix. In this framework, MSAs cluster together based on similarities in house price fluctuations. Shaded time periods indicate official NBER national recession dates.

Figure 3: Cluster Membership Based on Employment



This figure shows the posterior probability of cluster membership for each MSA in our sample using our baseline time-series clustering model with a similarity matrix. In this framework, MSAs cluster together based on similarities in employment growth fluctuations. Gray areas are not included in our sample.

Figure 4: Timing of Employment Downturns



This figure shows the posterior probability of regime timing for the aggregate variable z_t using our baseline time-series clustering model with a similarity matrix. In this framework, MSAs cluster together based on similarities in employment growth fluctuations. Shaded time periods indicate official NBER national recession dates.

Appendix: Estimation Details

In this section we outline the steps for the Gibbs sampler. The steps are quite similar to those outlined by Hamilton and Owyang (2012), but we incorporate an extension to the more general case of including a similarity weighting matrix. Additionally, the sampler for the cluster association matrix H differs from Hamilton and Owyang (2012) in that each entity is only allowed to be a member of one cluster whereas they allow for overlapping clusters.

We partition the parameters and latent variables into six blocks. Each block is drawn from its conditional distribution given the other blocks. The six steps are as follows:

1. Draw $\mu|Y,W,\sigma,\rho,H,Z$
2. Draw $\sigma|Y,W,\mu,\rho,H,Z$
3. Draw $\rho|Y,W,\mu,\sigma,H,Z$
4. Draw $H|Y,W,\mu,\sigma,\rho,Z$
5. Draw $P|Z$
6. Draw $Z|Y,W,\mu,\sigma,\rho,H,P$

We define $\hat{y}_t = [I-\rho W]^{-1} y_t$.

1. Draw $\boldsymbol{\mu|Y,W,\sigma,\rho,H,Z}$

We assume a normal prior for the growth rate parameters $\mu_n = [\mu_{n0}, \mu_{n1}]'$:

$\mu_n \sim N(m_{n0}, \sigma_n^2 M_{n0})$.

The posterior distribution is then given by $\mu_n \sim N(m_{n1}, \sigma_n^2 M_{n1})$ where:

$m_{n1} = M_{n1}\left(M_{n0}^{-1} m_{n0} + X_n' \widehat{Y_n}\right)$,

$M_{n1} = (M_{n0}^{-1} + X_n' X_n)^{-1}$,

$X_n = [X_{n1}', \ldots, X_{nT}']'$, $X_{nt} = [1 \; h_n(z_t)]'$, and $\hat{Y}_n = [\hat{y}_{1n}, \ldots, \hat{y}_{Tn}]'$

2. Draw $\boldsymbol{\sigma|Y,W,\mu,\rho,H,Z}$

We assume an inverse-gamma prior for each entity's variance parameter:

$$\sigma_n^{-2} \sim \Gamma\left(\frac{\upsilon_0}{2}, \frac{\tau_0}{2}\right)$$

The posterior distribution for σ_n^2 is given by:

$$\sigma_n^{-2} \sim \Gamma\left(\frac{\upsilon_0 + T}{2}, \frac{\tau_0 + \tau_1}{2}\right),$$

where:

$$\tau_1 = \sum_{i=1}^{T} [\hat{y}_{tn} - \mu_n{}'\mathrm{X}_{nt}]^2$$

3. Draw $\rho|Y, W, \mu, \sigma, H, Z$

We define $\tilde{y} = [\tilde{y}_1', \dots, \tilde{y}_T{}']'$ where:

$$\tilde{y}_t = [\tilde{y}_{t1}, \dots, \tilde{y}_{tn}]'$$

And:

$$\tilde{y}_{tn} = \frac{y_{tn} - \mu_n{}' X_{nt}}{\sigma_n}$$

Additionally, we define:

$$\ddot{X} = \begin{bmatrix} \rho W \ddot{y}_1 \\ \vdots \\ \rho W \ddot{y}_T \end{bmatrix}$$

Where:

$$\ddot{y}_t = [\ddot{y}_{t1}, \dots, \ddot{y}_{tN}]'$$

And:

$$\ddot{y}_{tn} = \frac{y_{tn}}{\sigma_n}$$

Assuming a normal prior for ρ:

$$\rho \sim N(r_0, R_0),$$

The posterior distribution is:

$$\rho \sim N(r_1, R_1),$$

Where:

$$r_1 = R_1(R_0^{-1} r_0 + \ddot{X}'\tilde{y}),$$

$$R_1 = (R_0^{-1} + \ddot{x}'\ddot{x})^{-1}$$

4. Draw $H|Y, W, \mu, \sigma, \rho, Z$

The cluster membership indicators h are drawn entity-by-entity. In contrast to Hamilton and Owyang (2012), we restrict each MSA to be a member of only one cluster as in Hernández-Murillo et al. (2017) and Francis et al. (2019).

We first calculate the conditional likelihood for country n to be a member of each cluster $k = 1, \dots, K$:

$$p(Y_n | h_{nk} = 1, W, \mu_n, \sigma_n, \rho, Z)$$

We then combine this conditional likelihood with a prior distribution $p(h_{nk} = 1)$ to get the posterior:

$$P_r(h_{nk} = 1|Y, W, \mu_n, \sigma_n, \rho, Z) = p(Y_n|h_{nk} = 1, W, \mu_n, \sigma_n, \rho, Z)p(h_{nk} = 1)$$

$$\sum_{j=1}^{K} p\big(Y_n\big|h_{nj} = 1, W, \mu_n, \sigma_n, \rho, Z\big)p(h_{nj} = 1)$$

We assume a uniform prior distribution for cluster membership: $p(h_{nk} = 1) = \frac{1}{K}$

5. Draw $P|Z$

We draw the transition matrix P similar to the step outlined by Hamilton and Owyang (2012). Since the columns P_i of P are independent, we assume a Dirichlet prior distribution for P_i:

$$P_i \sim D(p_{1i}, p_{2i}, \dots, p_{K+2i})$$

With restrictions to ensure zero transition probability between idiosyncratic regimes.

Thus, the posterior distribution for P_i is given by:

$$P_i \sim D(p_{1i} + N_{1i}(Z), p_{2i} + N_{2i}(Z), \dots, p_{K+2i} + N_{K+2i}(Z))$$

Where N_{ji} counts the number of transitions in Z from regime i to regime j.

6. Draw $Z|Y, W, \mu, \sigma, \rho, H, P$

Similar to Francis et al. (2019), we use the multi-regime filter outlined by Hamilton (1989). We first calculate the filter density forward for $t = 1, \dots, T$: $p(Z_t|Y_t, W, \mu, \sigma, \rho, H, P)$.

We then draw $Z_{t-1}, \dots, Z$, recursively by updating the forward filter densities:

$$p(Z_t|Z_{t+1}, Y_T, W, \mu, \sigma, \rho, H, P) = \frac{PZ_{t+1}Z_t\, p(Z_t|Y_t, W, \mu, \sigma, \rho, H, P)}{\sum_{k=1}^{K+2} PZ_{t+1}k\, p(Z_t = k|Y_t, W, \mu, \sigma, \rho, H, P)}$$

Where p_{ji} are the transition probabilities from P.